

代价敏感深度学习方法研究综述

吴雨茜 王俊丽 杨 丽 余森森

(同济大学电子与信息工程学院 上海 201804)

摘 要 代价敏感(cost-sensitive)学习方法能够有效缓解分类任务中的数据不平衡问题,已经被成功应用于各种传统机器学习技术中。随着深度学习技术的不断发展,代价敏感方法重新成为了研究热点。将深度学习与代价敏感方法相结合,不仅能够突破传统机器学习技术的限制,同时能够提高模型对数据的敏感性和分类的准确性,尤其是当数据中存在一定的不平衡性时。然而,如何有效地将两者进行结合成为了研究的重点和难点。研究学者从网络结构、损失函数和训练方法等多方面入手,不断提高深度学习结合代价敏感方法模型的性能。文中针对深度学习与代价敏感方法相结合的发展历程进行详细阐述,对几种具有创新性的模型进行了分析,并对比了模型的分类性能,最后对深度学习与代价敏感方法相结合的发展趋势进行了探讨。

关键词 代价敏感,深度学习,不平衡数据,神经网络

中图法分类号 TP18 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.05.001

Survey on Cost-sensitive Deep Learning Methods

WU Yu-xi WANG Jun-li YANG Li YU Miao-miao

(College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

Abstract Cost-sensitive learning method can effectively alleviate the problem of data imbalance in classification tasks and has been successfully applied to various traditional machine learning techniques. With the continuous development of deep learning technology, cost-sensitive method has become a research hotspot again. The combination of deep learning with cost-sensitive methods can not only breaks through the limitations of traditional machine learning technology, but also improve the data sensitivity and classification accuracy of the model, especially when there is a certain imbalance in the data. However, how to effectively combine the above two factors has become the focus and difficulty of the research. From the aspects of network structure, loss function and training method, researchers have improved the performance of the deep learning model combined with cost-sensitive method. In this paper, the development of the combination of deep learning and cost-sensitive method was described in detail, several innovative models were analyzed and the classification performance of these model was compared. Finally, the development trend of combination of deep learning and cost-sensitive method was discussed.

Keywords Cost-sensitive, Deep learning, Imbalanced data, Neural network

1 引言

分类是数据挖掘中的一项主要任务,而数据不平衡又是分类任务中的经典问题。不平衡数据集是指在一个数据集中,一个类的实例数很少,即少数类;而另一个类的实例数很多,即多数类^[1]。通常把这些数据集的分类问题称为不平衡数据分类,也称稀有类分类。在大部分的分类问题中,各类别的数据数量不可能完全相等,不同的差异程度对分类准确度的影响也有所不同。

数据不平衡问题广泛存在于实际应用领域,如网络入侵、临床诊断、信用卡欺诈等。在医疗诊断中,把一个病人误诊为

正常人与把一个正常人误诊为病人所付出的代价是不同的,准确地识别出病人更为重要^[2]。而在欺诈交易识别任务中,属于欺诈交易的应该是很少的,绝大部分的交易是正常的,这也是一个数据不平衡的分类问题^[3]。

目前,不平衡数据的研究越来越受到学者们的重视。针对数据不平衡问题的研究主要体现在两个方面:1)数据预处理方面,主要方法为抽样,通过修改训练集的样本分布来降低数据的不平衡程度^[4];2)算法层面,对现有分类算法进行改进,使之更倾向于少数类数据^[5]。其中,抽样又可以分为过采样(over-sampling)和欠采样(under-sampling)。过采样是指通过增加少数类样本来平衡数据集,其中 Chawla 等^[6]提出的

到稿日期:2018-06-17 返修日期:2018-09-27 本文受国家重点研发计划(2017YFB0304102),上海市科技创新行动计划(18511107400),同济大学中央高校基本科研业务费项目资助。

吴雨茜(1994—),女,硕士,主要研究方向为深度学习,E-mail:50827574@qq.com;王俊丽(1978—),女,博士,副研究员,主要研究方向为深度学习、网络数据分析、语义网络,E-mail:junliwang@tongji.edu.cn(通信作者);杨 丽(1992—),女,硕士,主要研究方向为深度学习;余森森(1989—),女,硕士,主要研究方向为概率主题模型。

SMOTE 方法为经典的代表性算法^[7]。该方法首先为每个稀有类样本随机选出几个邻近样本,并在该样本与这些邻近样本的连线上随机选取点,生成无重复的新的稀有类样本^[8]。欠采样则通过减少多数类样本来平衡数据集,例如由 Kubat 等^[9]提出的单边选择算法可以去除大类中的噪音样本、边界样本和冗余样本。虽然抽样的方法在一些数据集上取得了不错的效果,也能够与传统的机器学习方法(如决策树等)进行较好的结合^[10],但它们仍然存在很多缺陷:过采样由于只是重复地得到了一些人工生成的样本,会导致分类器产生过学习的现象,同时大大延长了训练时间;欠采样在去除多数类样本时,容易将重要的样本去除掉。因此,抽样方法调整不平衡数据的能力十分有限。

在算法层面,主要有集成学习、单类分类器和代价敏感等方法。集成学习的方法主要是将若干个基分类器组合成一个分类器,从而改善分类效果。常用的集成学习方法有随机森林(random forest)、装袋(bagging)、提升(boosting)等,其中 boosting 是应用最为广泛的集成技术。集成学习技术已经日趋完善,在线的 bagging 和 boosting 模型已经被应用于解决不平衡数据流的问题^[11]。AdaBoost 是采用提升技术的经典算法,在该算法中,分类器会提高分类错误样本的权值,降低分类正确样本的权值,最后得到由若干个基分类器组合而成的分类器。Chawla 等^[12]将 boosting 算法与 SMOTE 算法结合成 SMOTEBoost 算法,每次迭代时使用 SMOTE 来生成新的样本,以之取代 AdaBoost 算法原有的权值调整方法^[8]。在实际问题中,当只能获取单类样本集时,可以采用单类分类器的方法,只对一类数据进行训练。Abdi 等^[13]将采样与集成学习相结合,以达到优势互补的目的。

代价敏感是目前从算法层面解决不平衡数据问题的主要方法,它在解决分类问题时对于出现的不同分类错误使用不同的惩罚力度,其核心要素是代价矩阵。代价矩阵可被看作错误分类惩罚项的数字表示,因此代价敏感值的大小合适是该方法成功应用的关键^[14]。一般来说,代价敏感算法可以分为 3 类:1)改变原始的数据分布来得到代价敏感模型;2)对分类的结果进行调整,以达到最小损失的目的;3)直接构造一个代价敏感的学习模型。实验表明^[15-21],在很多应用领域,代价敏感的方法都优于其他方法。

近年来,随着深度学习方法的提出与发展,及其在解决图像、声音和文本分类问题上取得的令人瞩目的成绩,将代价敏感方法与深度学习方法相结合逐渐成为了学者们的研究热点。本文第 2 节介绍关于代价敏感的主要内容,包括其基本定义、代价敏感分类目标及研究现状;第 3 节详细介绍代价敏感与深度学习的结合应用及发展;第 4 节主要介绍几种有代表性的具有代价敏感性的深度学习模型,并对它们的分类效果进行对比和分析;第 5 节对未来代价敏感如何在深度学习中进行进一步应用展开探讨。

2 代价敏感

2.1 cost 定义

误分类代价有两种类型:基于类别的代价(class-depen-

dent cost)和基于样本的代价(example-dependent cost)。在实际应用中,基于类别的代价的应用范围更广。

代价敏感分类研究的是在具有不同 cost 值的数据集上的分类问题,通常使用 cost matrix 来描述数据集上的 cost 信息。

定义 1(代价矩阵, cost matrix) 给定数据集的类别数 N , 代价矩阵定义为一个 $N * N$ 的矩阵, 其中每一项 $c(i, j)$ 代表算法将 j 类的数据错误地分为 i 类所应当给予的惩罚值 $cost$, 原则上错误分类造成的 $cost$ 要大于正确分类造成的 $cost$ 。

2.2 代价敏感分类的目标

代价敏感分类的目标是学习一个分类器 $h: X \rightarrow Y$, 来对未知的样本进行分类, 使得期望代价最小:

$$\arg \min E_{(x, y)} \sim D[cost_{yh(x)}]$$

其中, D 表示样本的分布, E 表示期望代价, $h(x)$ 表示 x 的经验风险值, $cost_{yh(x)}$ 表示某个类别的分类代价值。

但由于样本分布 D 未知, 因此在训练过程中只能通过最小化经验风险得到 h :

$$\arg \min \sum_{i=1}^n cost_{y_i h(x_i)}$$

对于每个类别, 目标函数都尽可能使得所有类别的 $cost$ 值的概率之和最低。我们不再关注如何使得概率分布的值最大, 而是关注如何能够最小化分类结果对应的 $cost$ 值^[22]。这样虽然会在一定程度上放弃可能性最大的分类结果, 但在实际应用中有着重要的现实意义。

2.3 代价敏感分类算法的研究现状

代价敏感分类算法的研究主要体现在二分类问题上, 主要的实现方法分为两大类。一类是将任何以最小化错误率为目标的标准分类器转化为对代价敏感的分类器; 另一类是以代价为基准来得到不同的概率值, 试图将代价敏感性嵌入到算法中。

基于传统的学习模型, 根据错误分类的代价, 可以按照比例来变换训练集中类别的频率, 得到的代价敏感模型属于第一类方法。常用的手段为结合采样法来改变数据的类别分布。Cost-based sampling 方法由 Chan^[23]提出, 是代价敏感分类领域的开创性工作之一, 可以通过式(1)确定正样本的比例:

$$P(positive) = \frac{c(positive)i}{\sum_i c(positive)i} \quad (1)$$

其中, $c(positive)i$ 表示第 i 个正样本所对应的代价值。文献[23]同时提出了一种多类别的重采样方法来获取所需实例。

REBALANCE 方法由 Elkan^[24]提出, 同样是里程碑式的工作。Elkan 指出代价敏感分类问题的本质是优化问题, 通过调整训练数据集中数据实例的分布能够解决代价敏感分类问题。

通过对分类的结果进行后处理, 改变后验概率 $p(y|x)$ 的阈值, 使得代价高的样本不容易被分错, 这实际上是应用了贝叶斯风险最小化原理, 得到的代价敏感模型也属于第一类方法。MetaCost 是由 Domingos^[25]提出的一种算法, 通过对若干个原数据集进行重采样得到子集, 再组合若干个目标分类算

法,获得预测分类概率;然后,将训练集中的每一个实例重标记为代价最小的那个类别,重复运行目标分类算法,逐渐使得算法具有代价敏感的性质。MetaCost 采用的这种方法也被称作元代价,通过一个“元学习”的过程,修改训练样本的类标记,不断重新学习新的模型。此外,还可以通过投票的方式来组合弱分类器,并使用评价函数来调整各个弱分类器的权重,从而集成新的分类器来使模型具有代价敏感性,例如 Ada-Cost 算法^[26-28]。

嵌入式的方法需要对具体的算法进行修改和设计,已在一些常用的机器学习算法(如决策树、支持向量机(Support Vector Machine, SVM)、神经网络(Neural Network, NN)等)中取得了不错的效果。

Drummond 等^[29]提出了代价敏感的决策树,通过改变划分的标准,在传统的决策树算法中引入了误分类代价^[30]。实际上,修改剪枝策略也是使得决策树具有代价敏感性质的一种重要方法。传统的决策树根据错误率来进行剪枝,Knoll 等^[31]和 Bradford 等^[32]为决策树提出了代价敏感的剪枝方法,实现了在代价敏感的条件下对决策树进行剪枝,使得损失达到最小。郑恩辉等^[33]通过在 SVM 的设计中集成样本的不同误分类代价,得到代价敏感的 SVM。Adachi 等^[34]在此基础上又对代价敏感 SVM 进行了改进,将 SVM-score 作为标准来选取特征,提高了代价敏感 SVM 的准确率。对于神经网络,周宇航等^[35]改进了 SVM 的大间隔学习方法,得到了平均总代价更小的 CS-LDM 模型。Kukar 等^[36]对比了几种不同的优化途径,包括加入贝叶斯风险决策、改变输出层的代价敏感性、增加代价较高的数据的权重,以及修改目标函数,并通过提出新的后向传播算法改善了分类效果。Zhou 等^[37]把过采样、欠采样、阈值移动法、硬重组和软重组等基于采样的方法运用到 BP 神经网络,通过构造代价敏感的神经网络来解决不平衡数据的问题。

随着深度学习技术的不断发展,学者们又提出了可以将代价敏感与深度网络进行结合,以改善应用深度学习解决不平衡数据问题的现状。下面将详细介绍代价敏感与深度学习的结合应用及发展。

3 代价敏感与深度学习的结合

深度神经网络在处理特征提取、分类、识别以及预测等问题上取得了瞩目的成绩^[38-43],近几年有关代价敏感与深度学习结合的研究也越来越多。尤其是针对不平衡数据的分类问题,将代价敏感应用在深度神经网络(Deep Neural Network, DNN)中,使得模型在学习的过程中不仅能得到网络原本参数来最大化准确率,同时能够以总的代价最小化为目标来提高模型的性能,使深度学习能够更加灵活地应用在各种任务中。

本节将首先介绍深度学习的发展进程,接着主要从损失函数、代价矩阵、网络结构 3 方面来介绍具有代价敏感性的深度学习技术的发展情况。

3.1 深度学习的发展进程

深度学习也叫无监督特征学习,即无需人为设计特征提取,特征从数据中学习而来,实质上是多层表示学习方法的非

线性组合^[44]。深度结构非凸目标代价函数中普遍存在的局部最小,是训练困难性的主要来源。

深度学习隶属于神经网络体系。不同于浅层学习,深度学习能够使用更少的参数来逼近高度非线性函数。Hinton 等^[45-46]基于深信度网 DBN 提出非监督贪心逐层训练算法,并在随后提出多层自编码器深层结构^[45]。为了解决网络容易收敛到局部最小值的问题,Hinton 还提出了使用无监督预训练的方法来优化网络权值的初值,再进行权值微调,拉开了深度学习的序幕。

深度神经网络是由多个单层非线性网络叠加而成的^[47-48],编码器提供从输入到隐含特征空间的自底向上的映射,解码器以重建结果尽可能接近原始输入为目标将隐含特征映射到输入空间^[49]。深度神经网络可分为前馈深度网络(Feed Forward Deep Networks, FFDN)、反馈深度网络(Feed Back Deep Networks, FBDN)及双向深度网络(Bi-Directional Deep Networks, BDDN)3 种。

深度学习模型中的每个模块将低层的表示变换为高层的表示,逐层复合足够多的变换,以学习到非常复杂的函数^[50]。根据结构的不同,深度学习模型可分为深度神经网络 DNN、深度信任网络 DBN、栈式自动编码器(Stacked AutoEncoder, SAE)、深度玻尔兹曼机(Deep Boltzmann Machine, DBM)、卷积神经网络 CNN、递归神经网络(Recurrent Neural Network, RNN)等。其中,最具有代表性的网络结构为卷积神经网络和循环神经网络。

卷积神经网络是利用图像的局部特性训练数据,构造一个部分连通网络,隐层采用卷积核方式滑动计算。卷积神经网络具有局部感知、共享权重和池化降采样三大特点,被广泛应用于图像处理中。在处理不平衡图像数据的分类问题上,也常常使用由 CNN 构成的深度神经网络。Ren 等^[51]提出的 Faster R-CNN 是一个典型的基于深度学习的目标检测模型,分类准确度较高。Cheng 等^[52]为网络加入了优化模块 DCR,提高了模型的识别性能。CNN 在图像上的优异表现极大地拓展了其应用范围,Mahapatra 等^[53]将其应用于视网膜的分类问题上,Zhong 等^[54]提出 SatCNN 模型来解决卫星图像数据集的分类问题,以上研究均取得了不错的效果。对于数据量稀少或数据不平衡的情况,CNN 也能够较好地适应。Yue 等^[55]将 CNN 应用于解决不平衡的恶意软件图像分类问题,Windrim 等^[56]探索了在训练样本有限的情况下 CNN 的表现。

传统神经网络的输入和输出是相互独立的,而有些任务的后续输出与前阶段的特征是相关的,因此相关学者提出了具有“记忆”概念的递归神经网络 RNN^[50]。RNN 是深度模型中的隐层节点定向连接成环的人工神经网络。循环神经网络能够很好地对时间序列上的变化进行建模,主要应用于解决自然语言处理、语音识别、手写体识别等问题。为了解决循环神经网络面临的梯度消失或爆炸问题,Hochreiter 等^[57]提出长短时记忆(Long Short-Term Memory, LSTM),通过为每一个神经元引入门(gate)和存储单元,使得网络不仅具有记忆功能,同时能够处理较为复杂的序列数据。受人类注意力机制的启发,Mnih 等^[58]提出了注意力机制,该机制通过每次

处理注意力部分的数据,来减小任务的复杂度。

Shen 等^[59]使用基于 RNN 的网络来解决句子分类问题,Kang 等^[60]应用 RNN 网络来提高模型对演讲和音乐等语音数据的分类性能,并将其应用在助听器中。在情感分类任务中,RNN 的表现尤为优秀。Mun 等^[61]创新性地提出了基于 RNN 的特征提取方法,并将其用来进行情感分类。Cheng 等^[62]探究了 RNN 在中文微博文本的情感分类问题上的表现。以上研究均取得了一定的进展。

将 RNN 与 CNN 进行结合是近几年来研究热点,该类模型能够应用于多种问题上,并达到优势互补的效果。Wang 等^[63]提出了一个统一的多标签图片分类模型;Liu 等^[64]提出的 3DCNN-DQN-RNN 模型可以完成大规模 3D 点云的语义分析任务;Hori 等^[65]则选择了语音识别领域,使用联合模型来优化分类效果。此外,深入分析深度神经网络这个“黑匣子”也是研究学者一直努力的方向。熊红凯等^[66]从结构化、可解释化、多模态化 3 方面入手,剖析了深度神经网络的内在。

生成对抗网络(Generative Adversarial Networks, GAN)于 2014 年被 Goodfellow 等^[67]提出,并得到了广泛的关注。GAN 的核心思想来源于博弈论的纳什均衡,它由一个生成器(Generator)和判别器(Discriminator)构成,生成器的目的是尽量学习真实的数据分布,而判别器的目的是尽量正确判断输入数据是来自于真实数据还是生成器^[68]。生成对抗模型的训练是难点,因此基于 GAN 的衍生模型也被广泛提出。Arjovsky 等^[69]提出 W-GAN,用 Earth-Mover 来度量真实样本和生成样本分布之间的距离,用一个批评函数 f 来对应 GAN 的判别器,提升了 GAN 的训练效果。InfoGAN 是 GAN 的一个重要扩展,使得 GAN 能够学习到有效的语义特征,将输入噪声分为 z 和 c 两部分,将 c 作为隐码用于表征结构化隐层随机变量和具体特定语义之间的隐含关系^[70]。

将代价敏感与深度学习相结合来解决分类问题,能够进一步提升模型的性能,使其应用范围更为广泛。

3.2 损失函数

用损失函数来估量预测值与真实值的不一致程度,其值越小,模型的鲁棒性就越好。损失函数是经验风险函数的核心部分,也是结构风险函数的重要组成部分。常用的损失函数有对数损失(Logarithmic Loss)、平方损失(Square Loss)、指数损失(Exponential Loss)等。此外,新的损失函数也不断被提出。李跃波等^[71]分析了经典的平方损失函数,并提出了一种新的损失函数来进行质量的认证。Wang 等^[72]提出了一种在数据集不平衡的情况下有更好表现的损失函数,并使用该损失函数分别对正确分类和错误分类的误差进行计算。Cheng 等^[73]创新性地使用三联损失函数来提升卷积神经网络对多通道人脸图片的识别性能。

深度神经网络通过不断优化目标函数使得总的损失值最小来得到合适的网络权值和偏置值。通过改进损失函数可以使网络具有代价敏感性,常用的方式是添加 cost 参数 ω ,且对于不同的分类方式设置 ω 的不同取值。

Zhang 等^[74]使用深度代价敏感神经网络来处理语音分离问题,为原有的目标损失函数添加了新的权重 ω_s ,使其具

有代价敏感性。

$$\sum_{s=1}^S (\omega_s \min_a \sum_{m=1}^M l(x_{s,m}, x'_{s,m})) \quad (2)$$

其中, $l(*)$ 表示模型的损失函数, $x_{s,m}$ 表示模型的预测值, $x'_{s,m}$ 表示真实值, ω_s 的取值方法如下:

$$\omega_s = \frac{10^{-(\sigma_s)/20}}{\sum_{k=1}^S 10^{-(\sigma_k)/20}} \quad (3)$$

其中, $10^{-(\sigma_s)/20}$ 是语音和噪音的平均比值,不同语音数据的比值也不同,即相同分类结果下的 cost 不同,网络因此具有了代价敏感性。

Jiang 等^[75]使用深度代价敏感残差卷积神经网络(CS-ResCNN)来对精密的眼球图片进行分类,从而判断不同的眼科疾病。CS-ResCNN 改进了传统交叉熵(CE)损失函数的定义来使模型具有代价敏感性。

$$J(w) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k I\{y^{(i)} = j\} * C\{y^{(i)} = \min_class\} * \log \frac{e^{w_j^T x^{(i)}}}{\sum_{s=1}^k e^{w_s^T x^{(i)}}} \right] + \frac{\gamma}{2} \sum_{i=1}^k \sum_{j=1}^n w_{ij}^2 \quad (4)$$

其中, $C\{y^{(i)} = \min_class\}$ 是 cost 方程,如果 $y^{(i)}$ 是少数类标签,则值为 $C_{\min}$,否则值为 1。 $C_{\min}$ 则是通过训练学习模型得到。

Huang 等^[76]使用多个卷积神经网络(Convolutional Neural Network, CNN)并行来处理不平衡数据问题,同时从几何角度重新定义了类间的距离,提出了新的误差计算方式。Khan 等^[77]在改进损失函数方面做出了杰出的贡献,不同于其他仅仅添加 cost 因子的方法,他们提出的 CoSen 卷积神经网络同时对网络的参数和 cost 参数进行优化,并在多种经典损失函数(MSE loss, SVM loss, CE loss)上进行实验,提升了卷积神经网络的分类准确度。

3.3 网络结构

通过改变深度神经网络的网络结构来构造具有代价敏感性的模型具有很强的推广意义。Chung 和 Lin^[78-79]在 2015 年提出了代价敏感深度神经网络(CSDNN)模型,并于次年提出更加通用的内部辅助目标 AuxIT 框架。

CSDNN 模型通过 Pre-training 的方式使网络的输入具有代价敏感性,借鉴 Tu 和 Lin^[80]采用一对多(OVA)分解的方法来构建代价敏感支持向量机的思想,对输出层进行改良。

AuxIT 框架重复使用 CSDNN 模型在 Pre-training 过程中使用的代价敏感自动编码器(CSAE)结构,使网络的每一层都具有代价敏感性,同时在输出层采用更优的 ReLU 激活函数。这种方法可以使不同种类的深度神经网络具有代价敏感性。

3.4 代价矩阵

2.1 节介绍过代价矩阵(cost matrix),研究学者使用传统的机器学习方法来更新代价矩阵,使其能够更好地表示误分类的代价。

Zhang 等^[74]结合进化算法来获得更优的代价矩阵,提出使用差分进化算法的深度代价敏感置信网络(CSDBN-DE)来解决不平衡数据问题。首先,使用训练数据对 DBN 进行 Pre-training;接着,随机初始化代价矩阵,并根据训练误差来选择

更好的下一代数据,反复执行进化过程;最后,选择最佳的代价矩阵作为最终网络的代价矩阵。算法流程如算法 1 所示。

算法 1 结合进化算法的代价敏感算法

Pre-training 过程

1. 初始训练集为 S_{train} 。
2. 用逐层贪婪的方法来 Pre-training 深度信念网络 (Deep Belief Network, DBN)。

微调过程

1. 随机初始化代价矩阵作为群体。
2. 通过不同的操作进行突变和交叉来生成新的代价矩阵群体。
3. 将训练输出和误差的代价矩阵相乘。
4. 根据其表现,选出恰当的代价矩阵并丢弃较差的个体,从而进化成为新一代。
5. 迭代进行变异和选择的过程,直到代数达到最大。
6. 把找到的最好的代价矩阵加到 DBN 的输出层上,并使用测试集 S_{test} 进行测试。

Li 等^[81]结合三支决策来解决人脸图像的分类问题,由于 DNN 是双向决策模型,无法产生三支决策中的边界决策,因此他们改进了代价矩阵,将边界决策加入到矩阵中,如表 1 所列。需要注意的是,误分类的代价应大于边界分类的代价,边界分类的代价应大于正确分类的代价。即 $\omega_{PP} \leq \omega_{BP} \leq \omega_{NP}$, $\omega_{NN} \leq \omega_{BN} \leq \omega_{PN}$ 。

表 1 三支决策改进后的代价矩阵

Table 1 Improved cost matrix by three decisions

实际类别	正确决策	边界决策	错误决策
类别 A	ω_{PP}	ω_{BP}	ω_{NP}
类别 B	ω_{PN}	ω_{BN}	ω_{NN}

除了能够提高分类的准确度外,代价敏感方法还能用于加快深度神经网络的训练速度。Pang 等^[82]提出分层计算代价 (HCC) 的方法,并将其应用于卷积神经网络 CNN 中,找到有冗余参数的隐藏层,并使用不同的结构来平衡准确性和速度之间的关系。HCC 主要用于处理二分类问题,这里的代价指运算的时间代价。HCC 模型中,代价矩阵每一项的取值由特征点决定。如图 1 所示,假设一共有 20 个特征点,则代价矩阵为一个 20×20 的矩阵,假定特征点 19 由特征点 9,10,13,14 决定,则点对 (19,9), (19,10), (19,13), (19,14) 被赋值为 1,第 19 行的其余值为 0。

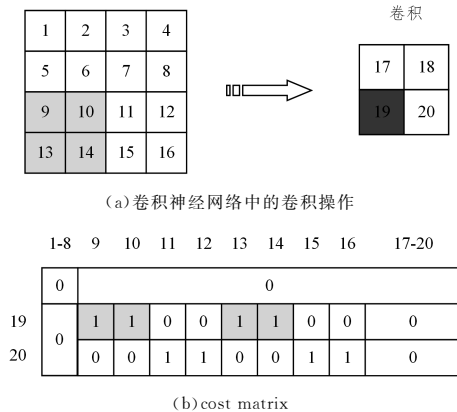


图 1 HCC 模型代价矩阵计算方式

Fig. 1 Calculation method of HCC model costing matrix

4 深度代价敏感的改进模型及其比较

3.1 节介绍了深度学习的发展历程,近年来不断有学者将深度学习与代价敏感学习方法进行结合,并取得了突破性的成绩。本节将重点介绍 3 个以深度神经网络为基础的结合代价敏感的分类模型。

4.1 CSDNN

CSDNN 是由 Chung 和 Lin^[78]提出的结合代价敏感和深度神经网络 DNN 的模型,也是第一个真正从算法本身出发进行改进的代价敏感深度学习算法。

CSDNN 模型改进了深度神经网络本身的训练方式,同时添加了 Pre-training 过程来得到具有代价敏感性的数据,使整个模型在训练过程中都能保持代价敏感性。

模型借鉴了 Tu 和 Lin 等^[80]采用一对多 (OVA) 分解的方法来构造代价敏感支持向量机的思想,将代价敏感分类问题分解为每一类别的回归问题,并使用回归因子对代价进行重新估算。

通过将深度神经网络的 softmax 层替换为回归输出,并使用以下公式来估计每一个输出神经元的代价,可以直接使用回归因子 r 代替代价来进行计算。

$$g_r(x) \equiv \arg \min_{1 \leq k \leq K} r_k(x) \quad (5)$$

其中, K 表示总的类别数, $r_k(x)$ 表示类别 k 中数据的回归因子。

深度神经网络采用反向传播算法来传递误差,由于 OVA 分解把 DNN 的损失改为了单边损失,因此需要一个平滑估计算法 smooth 来替代原来的误差传递方法。CSDNN 参考 Lee 和 Mangasarian^[83]提出的算法来定义平滑估计:

$$\delta_{n,k} \equiv \ln(1 + \exp(z_{n,k} * (r_k(x_n) - c_n[k]))) \quad (6)$$

其中, $r_k(x_n)$ 是数据 x_n 的回归因子, $c_n[k]$ 是其代价值, $\delta_{n,k}$ 可以近似看作模型的单边损失。

将该方法应用到 CSDNN 模型中,模型的 cost 损失就被重新定义为回归损失,其目的是使所有样本的损失和最小,即:

$$L_{\text{smooth}}(S) = \sum_{n=1}^N \sum_{k=1}^K \delta_{n,k} \quad (7)$$

Pre-training 在 2006 年由 Hinton 等^[84]提出,能够帮助神经网络更好地初始化权值,避免网络陷入局部极小中,同时能够缓解梯度消失问题。使用去噪自动编码器 (Denoising AutoEncoder, DAE) 是一种经典的 Pre-training 方法。DAE 由 Vincent 等^[85]提出,通过添加随机噪声的方式将原始的输入进行重编码。将多个 DAE 进行组合,能得到更深层的堆积去噪自动编码器 (Stack Denoising AutoEncoder, SDAE)。CSDNN 采用的就是改进的 DAE,使其具有代价敏感性,我们称之为代价敏感自动编码器 (Cost Sensitive AutoEncoder, CSAE)。

CSAE 在 DAE 的基础上增加了额外的参数,使得网络不仅能够为输入重新构建一个输出,同时添加了代价信息。最后损失函数变成了两部分损失的组合,可以使用参数 β 来进行平衡:

$$L_{CSAE}(S_c) = (1-\beta) * L_{CE}(S_c - \{c_n\}_{n=1}^N) + \beta * L_{SMOOTH}(S_c) \quad (8)$$

其中, β 用来平衡两部分损失。当 $\beta=0$ 时, CSAE 就退化为 DAE。这种改进可以使模型在 Pre-training 和训练两部分中都具有代价敏感性。CSDNN 模型的结构如图 2 所示。

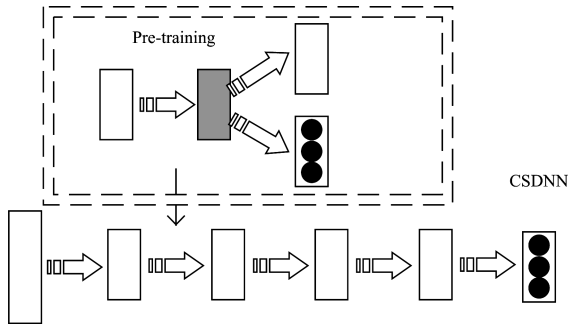


图 2 CSDNN 模型结构图

Fig. 2 Structure of CSDNN model

图 2 中的灰色框表示神经网络中的一个普通的隐藏层, 模型将普通的输入层扩展为两部分, 一部分负责训练和优化网络参数, 另一部共有 K 个节点, 负责训练和优化代价参数, 从而完成 Pre-training 的过程。

4.2 AuxIT

Chung 和 Lin^[79] 在 CSDNN 模型的基础上, 提出了一个更加通用的框架——内部辅助目标 (AuxIT), 其能够缓解原有的代价敏感模型在实际应用和网络深度上的限制。AuxIT 框架通过在每个隐藏层上添加额外的神经元, 得到逐层的代价估计值, 并把最小化代价的总和作为优化的目标。

CSDNN 模型仅仅改变了全连接层的激活函数, 并在数据输入网络之前先进行 Pre-training 来解决梯度消失等问题。然而, 随着新的激活函数如 ReLU 等的提出, 梯度消失问题也有了新的解决策略。因此, 在 AuxIT 框架中, 使用 ReLU 函数来代替 CSDNN 模型中的 Pre-training 过程。

在 CSDNN 模型中, 通过使用代价敏感自动编码器 CSAE 结构, 使输入的数据能够被重新编码, 同时具有代价敏感性。为了得到一个通用的框架, AuxIT 重复利用 CSAE 结构, 在网络的每一层中都加入 K 个额外的神经元, 得到的额外输出能够表示代价的估计值, 使得代价敏感能够贯穿于整个网络中。AuxIT 模型的结构如图 3 所示。

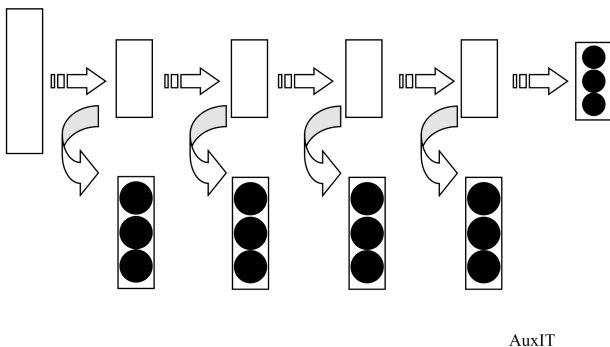


图 3 AuxIT 模型结构图

Fig. 3 Structure of AuxIT model

针对输出层, AuxIT 框架沿用了 CSDNN 模型中的改进方式, 使得输出层也具有代价敏感性。最后需要解决的是损失函数的问题, 模型采用权值混合损失函数, 网络本身的单边回归损失函数与额外神经元的单边回归损失函数通过用权值来平衡的方式进行组合。计算模型损失函数的方式如下:

$$\sum_{i=1}^{H-1} \alpha_i * L_{OSR}^{(i)} + L_{OSR}^{(*)} \quad (9)$$

其中, H 表示隐层的编号, α_i 是一个平衡因子。AuxIT 框架能够很好地被应用在各种深度网络中, 如应用于图像处理的 CNN 等, 也能够解决一些大规模的代价敏感分类问题。

4.3 CoSen CNN

除了对网络结构进行优化外, 学者们还对如何解决代价敏感分类问题中新引入参数的更新问题进行了研究。Khan 等^[77] 提出了一种能够自动学习数据特征表示的代价敏感深度神经网络, 其能够很好地解决二分类问题和多分类问题, 具有较强的鲁棒性。通过对代价参数和网络本身的参数进行联合优化来对模型进行训练, 采用“保持一个参数不变, 训练另一个参数”的方式完成对两部分参数的共同优化。此外, 这种训练方式对于不同的 loss 函数都能适用。

模型采用以下代价敏感误差函数来定义训练数据的平均 loss:

$$E(\theta, \varphi) = \frac{1}{M} \sum_{i=1}^M l(d^{(i)}, y_{\theta, \varphi}^{(i)}) \quad (10)$$

其中, θ 是网络本身的参数 (包括权值 w 和偏置 b), φ 是代价的参数, M 代表训练样本的总数, $d^{(i)}$ 是网络的期望输出, $y_{\theta, \varphi}^{(i)}$ 是网络的实际输出。 $l(*)$ 表示损失函数, 可以是均方差 (Mean-Square Error, MSE) 损失、支持向量机 (SVM) 折叶损失, 或是交叉熵 (Cross-Entropy, CE) 损失。表 2 列出了如何构造具有代价敏感性的损失函数, 其中 ϵ 代表 cost 矩阵。在 cost 矩阵中, 每一个类别对应其他的类别都有一个确定的关系。 ϵ 和 φ 的关系可以通过以下公式确定:

$$\epsilon(p, q) = \begin{cases} \max(\varphi_p, \varphi_q), & p \neq q, (p, q) \in c \\ \varphi_p, & p = q, p \in c \end{cases}$$

模型的优化目标就是找到最优的参数 (θ^*, φ^*) , 能够使得的 cost 最小, 即:

$$(\theta^*, \varphi^*) = \arg \min_{\theta, \varphi} E(\theta, \varphi) \quad (11)$$

表 2 3 种不同的代价敏感损失函数

Table 2 Three different cost-sensitive loss functions

	损失 loss	代价敏感损失 loss
MSE	$l(d^{(i)}, y^{(i)}) = \frac{1}{2} \sum_n (d_n^{(i)} - y_n^{(i)})^2$	$y_n^{(i)} = \frac{1}{1 + \exp(-o_n^{(i)} \epsilon(d_n^{(i)}))}$
SVM	$l(d^{(i)}, y^{(i)}) = -\sum_n \max(0, 1 - (2d_n^{(i)} - 1)y_n^{(i)})$	$y_n^{(i)} = o_n^{(i)} \epsilon(d_n^{(i)})$
CE	$l(d^{(i)}, y^{(i)}) = -\sum_n (d_n^{(i)} \log y_n^{(i)})$	$y_n^{(i)} = \frac{\epsilon(d_n^{(i)}) \exp(o_n^{(i)})}{\sum_k \epsilon(d_n^{(i)}) \exp(o_k^{(i)})}$

如何有效地对两部分参数进行优化呢? CoSen CNN 模型的参数优化方式参考了 Raj 等^[86] 提出的针对不平衡数据的卷积神经网络优化算法: 固定代价参数 φ 不变, 更新网络参数 θ 。每一轮数据被训练完后, 在当前网络参数 θ 不变的情况

下,更新 cost 参数 φ 。算法流程如算法 2 所示。

算法 2 针对卷积神经网络的代价敏感算法

输入:训练集 (x, d) , 验证集 (x_v, d_v) , 最大训练轮次 (M_{ep}) , θ 的学习率 γ_θ , ξ 的学习率 γ_ξ 。

输出:训练好的参数 (θ^*, ξ^*) 。

训练过程:

1. 建立卷积神经网络 CNN。
 2. 随机初始化网络参数 θ 。
 3. 将 ξ 和误差 err 赋值为 1。
 4. 在每一轮次训练开始后,计算 ξ 的梯度,更新代价矩阵的参数 ξ^* , 并把 ξ^* 赋值给 ξ 。
 5. 在每一个 batch 中,利用前向传播计算出网络输出 out,利用后向传播计算出梯度 grad,并更新参数 θ^* ,最后将其赋值给 θ 。
 6. 每一 batch 训练结束后,利用前向传播计算出误差 err*, 如果新的误差比原来的误差大,就降低 γ_ξ ,并将新的误差值赋给 err。
- 所有轮次训练结束后,输出训练好的参数 (θ^*, ξ^*) 。

神经网络采用随机梯度下降法 (Stochastic Gradient Descent, SGD) 进行误差的反向传播,因此可以直接对网络本身的参数 θ 进行优化。对于代价参数 φ ,模型使用以下公式来定义代价函数:

$$F(\varphi) = \frac{1}{2} \sum_n (h_n - \varphi_n)^2, n \in [1, C] \quad (12)$$

其中, C 表示类别数, h 表示训练数据类分布的向量表示。网络的最终目的是得到最优的 φ^* 以使 $F(\varphi)$ 的值最小,即:

$$\varphi^* = \arg \min_{\varphi} F(\varphi) \quad (13)$$

在反向传播的过程中,具有代价敏感性质的 loss 的定义不同,梯度的计算方式也有所不同,表 3 列出了 MSE, SVM, CE 3 种损失函数的梯度计算表达式, $II(\cdot)$ 表示一个指标函数。可以发现,虽然加入了代价敏感参数,CE loss 的梯度计算方式并没有发生改变。

表 3 3 种损失函数的梯度计算表达式

Table 3 Gradient calculation expressions of three loss functions	
	偏导 derivative
MSE	$\frac{\partial l(d^{(i)}, y^{(i)})}{\partial \theta_n^{(i)}} = -\epsilon(d_n^{(i)}) (d_n^{(i)} - y_n^{(i)}) y_n^{(i)} (1 - y_n^{(i)})$
SVM	$\frac{\partial l(d^{(i)}, y^{(i)})}{\partial \theta_n^{(i)}} = -(2d_n^{(i)} - 1) \epsilon(d_n^{(i)}) II\{1 > y_n^{(i)} \epsilon(d_n^{(i)}) (2d_n^{(i)} - 1)\}$
CE	$\frac{\partial l(d^{(i)}, y^{(i)})}{\partial \theta_n^{(i)}} = -d_n^{(i)} + y_n^{(i)}$

CoSen CNN 所采用的网络结构为一个深度为 18 层的卷积神经网络,具体结构参考了 Simonyan 和 Zisserman^[87]提出的一种卷积神经网络结构,只是在输出层前添加了两个额外的全连接层。模型采用的损失层具有代价敏感性,因此含有大量的训练参数。为了达到使用少量的训练数据就能得到一个较优的训练结果的目的,模型首先采用 Simonyan 和 Zisserman 提出的训练方法对前 16 层网络进行 Pre-training,然后设定一个较高的学习率来对初始权值随机的后两层网络进行训练。

4.4 3 种改进模型的效果比较

将这 3 种改进模型应用在不平衡数据的分类上,可以衡

量它们在分类问题上的表现。其中主要采用手写数字分类数据集 (MNIST) 和 CIFAR 图像数据集 (CIFAR-10 和 CIFAR-100) 3 个库来进行比较和评价。除此之外,不同模型采用的网络结构不同,所针对的数据类型也大不相同。以不平衡度作为度量标准,可以得到对 3 种模型分类效果更加直观的认识。

1) 手写数字分类数据集: MNIST

MNIST 数据集包含有 70 000 张手写图片,分别为数字 0—9 中的一类。

2) CIFAR 图像数据集: CIFAR-10, CIFAR-100

CIFAR 数据集被用来识别物体,包含 60 000 张图片。CIFAR-10 和 CIFAR-100 唯一的区别在于类别数不同。

由于 MNIST 数据集和 CIFAR 数据集都是平衡数据集,而代价敏感模型主要用于处理不平衡数据问题,因此需要对原始数据集进行调整。

Chung 和 Lin^[78-79]随机地选择了所有类别中的 40%,并移除了这些类别中 70% 的样本,得到了实验所采用的数据集 $MNIST_{imb}$ 、 $CIFAR-10_{imb}$ 以及 $CIFAR-100_{imb}$ 。

Khan 等^[77]则将数据集中所有的奇数或偶数类的样本减少 10%~15%,使数据具有不平衡性,并在不平衡度不同的数据集上进行实验。

本小节主要从代价值和准确度两个角度来比较和分析 3 种改进模型分类效果。

1) 代价值 (cost value)

从代价值最小化的角度,可以对 CSDNN 模型和 AuxIT 框架的性能进行比较^[78-79]。由于 CSDNN 模型存在局限,为了保证实验的公平性,采用层数相同的全连接网络结构和同样的激活函数 ReLU。AuxIT 框架需要一个恰当的参数来协调两种损失函数之间的平衡,这里取通过实验得到的最佳值 0.2^[79]。同时,实验中还将没有加入任何代价敏感优化的网络结构 NaiveDNN 一同进行比较。实验结果如图 4—图 6 所示,其中横坐标表示隐藏层个数。

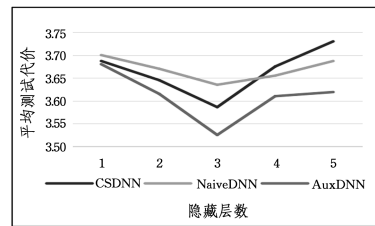


图 4 不同网络的平均测试代价曲线 (CIFAR-10_{imb} 数据集)

Fig. 4 Average test cost curves for different networks (CIFAR-10_{imb} dataset)

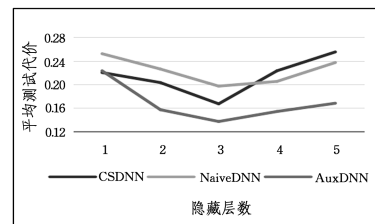
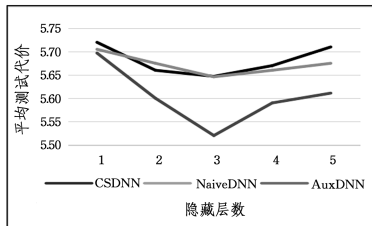


图 5 不同网络的平均测试代价曲线 (MNIST_{imb} 数据集)

Fig. 5 Average test cost curves for different networks (MNIST_{imb} dataset)

图 6 不同网络的平均测试代价曲线(CIFAR-100_{imb}数据集)Fig. 6 Average test cost curves for different networks
(CIFAR-100_{imb} dataset)

可以看到,随着网络层数的不断加深,CSDNN 模型发生了梯度消失,测试时随着层数的加深,总代价值的增长率越来越大,甚至超过了没有加入代价敏感优化的 NaiveDNN;但使用 AuxIT 框架的网络的实验效果十分稳定,测试时的总代价值远远小于 CSDNN 和 NaiveDNN。当网络层数为 3 时,3 种网络均得到了最小的代价值。

相比于 CSDNN 模型,AuxIT 框架具有使用灵活、方便的优点,可适用于各种不同的网络类型,更具有推广价值。

2) 准确度(accuracy)

CoSen CNN 模型在分类精确度上取得了良好的效果,大大提高了对不平衡的 MNIST 和 CIFAR-100 数据集的分类能力。使用没有加入代价敏感损失层的卷积神经网络作为对照网络 baseline CNN,对比两种网络在不平衡的 MNIST 和 CIFAR-100 数据集上的表现,结果如表 4 和表 5 所列^[77]。

表 4 不同网络的准确率比较(MNIST 数据集)

Table 4 Accuracy comparison of different networks
(MNIST dataset)

(单位: %)		
测试集	Baseline	CoSen
将数据集中偶数部分减少 10%	97.6	98.5
将数据集中奇数部分减少 10%	97.1	98.3
将数据集中偶数部分减少 25%	98.1	98.9
将数据集中奇数部分减少 25%	97.8	98.5

表 5 不同网络的准确率比较(CIFAR-100 数据集)

Table 5 Accuracy comparison of different networks
(CIFAR-100 dataset)

(单位: %)		
测试集	Baseline	CoSen
将数据集中偶数部分减少 10%	55.0	60.1
将数据集中奇数部分减少 10%	53.8	59.7
将数据集中偶数部分减少 25%	57.7	61.5
将数据集中奇数部分减少 25%	57.4	61.6

可以看出,对于 MNIST 数据集,将偶数类中的数据移除能够带来更好的实验效果。随着数据集不平衡度的增加,模型的效果反而更好,说明模型能够很好地解决数据不平衡问题。为了更具针对性,实验进一步对比了两种网络在 MIT-67 和 Caltech-101 数据集上的表现,结果如表 6 和表 7 所列。其中,MIT-67 是一个包含 15620 张图片的数据集,共被分为 67 类;Caltech-101 是一个包含 9144 张图片的数据集,共被分为 102 类。这两个数据集的不平衡度更大,其实验结果也更有说服力。

MIT-67 数据集的不平衡度约为 7, Caltech-101 数据集的不平衡度约为 26, 加入了代价敏感的 CNN 在分类准确率上有了显著的提升。

对比 3 种改进模型可以看出,对于结构较为简单的深度神经网络,CSDNN 模型的改进方法更加便于操作,其分类表现也较好。通过在预训练阶段增加数据的代价敏感性,能够在一定程度上提升模型的效果。但是,CSDNN 模型只能用于结构单一的全连接网络,当网络结构较为复杂(例如卷积神经网络 CNN 等)时,使用较为通用的 AuxIT 框架更为恰当。AuxIT 框架借鉴 CSDNN 模型在预训练时的改进方法,通过改变深度神经网络中每一层的节点组成,有针对性地为网络添加代价敏感性。而对于代价矩阵不明确的情况,相比于以上两种模型,使用 CoSen CNN 模型中提出的同步训练的方法更加合适。同步训练不仅能够增加模型的代价敏感性,进而提高分类的准确性,而且对于确定代价矩阵 cost matrix 的初值也有着重要作用。此外,同步训练的方法对于网络结构没有过多的要求,也能够广泛地适用于各种不同的损失函数。

表 6 不同网络的准确率比较(Caltech-101 数据集)

Table 6 Comparison of accuracy of different networks
(Caltech-101 dataset)

(单位: %)		
测试集	Baseline	CoSen
将数据集按照 60%:40% 划分	88.1	89.2
将数据集中偶数部分减少 10%	77.4	83.0
将数据集中奇数部分减少 10%	76.1	82.2
将数据集按照 30%:70% 划分	85.5	87.9
将数据集中偶数部分减少 10%	74.6	80.3
将数据集中奇数部分减少 10%	75.2	80.9

表 7 不同网络的准确率比较(MIT-67 数据集)

Table 7 Accuracy comparison of different networks
(MIT-67 dataset)

(单位: %)		
测试集	Baseline	CoSen
将数据集按照 60%:40% 划分	70.7	73.2
将数据集中偶数部分减少 10%	50.4	57.0
将数据集中奇数部分减少 10%	50.1	56.4
将数据集按照 30%:70% 划分	61.9	66.3
将数据集中偶数部分减少 10%	38.7	44.8
将数据集中奇数部分减少 10%	37.2	43.5

通过对数据的不平衡度进行分析可知,CSDNN 模型和 AuxIT 框架移除了 70% 的样本,其数据的不平衡度更大,实验也更加具有挑战性,说明这两种模型广泛适用于各种场景,能够处理较多的不平衡分类问题。实验时,CoSen CNN 模型采用将平衡数据减少 10%~15% 的方法来达到样本不平衡的目的,因此其数据的不平衡性较弱。

5 深度代价敏感模型的发展趋势

深度学习技术发展到今天,越来越多地被应用于解决实际问题。不平衡数据问题切合实际并且亟待解决,将代价敏感算法与深度神经网络相结合,能够有效地改善 DNN 在解决不平衡数据问题上的研究现状。在未来,深度代价敏感模型会在更多的应用领域发挥重要的作用。

5.1 多样化神经网络结构

代价敏感神经网络从简单的人工神经网络发展到多层的深度神经网络,再到一些复杂的网络结构,如卷积神经网络 CNN、残差网络 ResNet 等,其网络结构呈现多样化的趋势。

Zhang 等^[74]使用深度置信网络 DBN 作为基本网络结构,在 KEEL 数据集上取得了不错的效果。Khan 等^[77]在 Simonyan 和 Zisserman 提出的深度达到 16 层的卷积神经网络的基础上添加了全连接层和损失层,最终的网络为一个 18 层的 CNN。Jiang 等^[75]为更为复杂的残差卷积神经网络 (ResCNN) 加入了代价敏感性,提高了网络对精密的眼球图片的分类准确率。Chung 和 Lin^[79]则使用了经典的深度神经网络结构 VGG 和 ResNet,提高了算法的普适性。

在目前的工作中,与循环神经网络 RNN 结合的深度代价敏感模型非常稀少。RNN 能够有效地处理和预测序列数据,如果能够在其中加入代价敏感性,将会提高深度学习技术在多个应用领域解决问题的能力。此外,生成式对抗网络 GAN 的提出引发了学者们的研究热潮,GAN 不仅在无监督学习上表现出优秀的能力,其变形在半监督学习上的表现也同样出色。Fiore 等^[88]已经尝试使用 GAN 来生成不平衡数据中的少数类数据,并将 GAN 应用于检测信用卡欺诈,取得了一定的效果。未来,深度代价敏感模型的网络结构也会越来越多样化和复杂化。

5.2 紧密联系传统机器学习算法

代价敏感在传统的机器学习算法上已经取得了杰出的效果,未来可以结合现有的研究成果,如代价敏感决策树、代价敏感支持向量机、代价敏感随机森林等,使代价敏感能够更好地融合到深度神经网络中。

Chung 和 Lin 等^[78]在处理 DNN 的输出层上所采用的解决方法就借鉴了前人构造代价敏感支持向量机的思路,且取得了不错的效果。Li 等^[81]结合三支决策来解决人脸图像的分类问题,修改了原有的代价矩阵,使得分类过程更加合理。Zhang 等^[74]结合差分进化算法,尝试获得一个最优的 cost matrix,提高了分类的准确性。

结合已经成熟的代价敏感机器学习模型,可以在一定程度上对代价矩阵进行优化,也可以为如何使模型具有代价敏感性提供解决思路。

5.3 创新参数训练方式

现有的训练方式大多是将代价作为权值因子加入到损失函数中,在代价矩阵已确定好的情况下,对网络进行迭代和更新。这种训练方式存在一定的局限,无法获得一个最优的代价矩阵。

Khan 等^[77]创新性地提出了同时对网络本身的参数和代价敏感参数进行更新的训练方式,采用“保持一个参数不变,训练另一个参数”的方法来获取代价矩阵,取得了优秀的分类效果。Geng 等^[89]在此基础上加入了对时间序列的考虑,扩大了模型的应用范围。Chung 和 Lin^[78]采用 Pre-training 的方式首先对训练数据进行编码处理,使数据具有代价敏感性,然后将处理后的数据传入到深度神经网络中进行训练,能够有效地提高分类结果的准确率。

创新参数的训练方式对于获得一个合理的代价值有着重要的意义,同时能够缩短训练所需的时间,提高模型的准确率。

结束语 本文对代价敏感算法模型展开讨论,进行了全面的综述;分析了代价敏感对于处理不平衡数据问题的重要意义,从网络结构、损失函数、代价矩阵几个方面探讨了深度代价敏感模型的研究现状;对深度代价敏感模型的几种改进方法进行了详细的探究与比较;最后对深度代价敏感模型的未来发展方向进行了展望。

目前,代价敏感已经成为改进深度学习技术的一个重要切入点,能够有效地提高分类的准确度,并且能够广泛地应用在多种实际问题的处理中。最近,有越来越多的学者在研究深度代价敏感模型相关的问题,通过丰富网络结构、结合传统机器学习方法以及创新训练方式等途径提升模型的性能,这也是我们未来努力的方向。

参考文献

- [1] SUN Y, WONG A K C, KAMEL M S. Classification of Imbalanced Data: A REVIEW[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2009, 23(4): 687-719.
- [2] KHALILIA M, CHAKRABORTY S, POPESCU M. Predicting disease risks from highly imbalanced data using random forest [J]. BMC Medical Informatics & Decision Making, 2011, 11(1): 51.
- [3] ZHANG D, XU W. A Data-Distribution-Based Imbalanced Data Classification Method for Credit Scoring Using Neural Networks [C] // Sixth International Conference on Business Intelligence and Financial Engineering. IEEE, 2014: 557-561.
- [4] XIONG B Y, WANG G Y, DENG W B, et al. Unsaturated data undersampling method based on sample weight [J]. Journal of Computer Research and Development, 2016, 53(11): 2613-2622. (in Chinese)
熊冰妍, 王国胤, 邓维斌, 等. 基于样本权重的不平衡数据欠抽样方法[J]. 计算机研究与发展, 2016, 53(11): 2613-2622.
- [5] FERNÁNDEZ A, LÓPEZ V, GALAR M, et al. Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches [J]. Knowledge-Based Systems, 2013, 42(2): 97-110.
- [6] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique [J]. Journal of Artificial Intelligence Research, 2002, 16(1): 321-357.
- [7] CHAWLA N V. Data Mining for Imbalanced Datasets: An Overview [M] // Data Mining and Knowledge Discovery Handbook. US: Springer, 2005: 853-867.
- [8] YE Z F, WEN Y M, LU B L, et al. A Review of Researches on Unbalanced Classification [J]. Journal of Intelligent Systems, 2009, 4(2): 148-156. (in Chinese)
叶志飞, 文益民, 吕宝粮, 等. 不平衡分类问题研究综述[J]. 智能系统学报, 2009, 4(2): 148-156.
- [9] KUBAT M, MATWIN S. Addressing the Curse of Imbalanced Training Sets: One-Sided Selection [C] // International Conference on Machine Learning. 1997: 179-186.

- [10] SARDARI S, EFTEKHARI M. A Fuzzy Decision tree approach for imbalanced data classification[C]// International Conference on Computer and Knowledge Engineering. IEEE, 2017: 292-297.
- [11] WANG B, PINEAU J. Online Bagging and Boosting for Imbalanced Data Streams[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(12): 3353-3366.
- [12] CHAWLA N V, LAZAREVIC A, HALL L O, et al. SMOTE-Boost: Improving Prediction of the Minority Class in Boosting [C]// Knowledge Discovery in Databases: Pkdd 2003, European Conference on Principles and Practice of Knowledge Discovery in Databases, Cavtat-Dubrovnik, Croatia, DBLP, 2003: 107-119.
- [13] ABDI L, HASHEMI S. To combat multi-class imbalanced problems by means of over-sampling and boosting techniques[M]// Soft Computing. Springer-Verlag, 2015: 3369-3385.
- [14] TAO X M, HAO S Y, ZHANG D X, et al. An Overview of Unbalanced Data Classification Algorithms [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science), 2013, 25(1): 101-110. (in Chinese)
陶新民, 郝思媛, 张冬雪, 等. 不均衡数据分类算法的综述[J]. 重庆邮电大学学报(自然科学版), 2013, 25(1): 101-110.
- [15] YANG J M, GAO C, QU C Y, et al. Cost-sensitive Random Forest Unbalanced Data Classification Algorithm [J]. Science Technology and Engineering, 2018(6): 285-290. (in Chinese)
杨杰明, 高聪, 曲朝阳, 等. 基于代价敏感的随机森林不平衡数据分类算法[J]. 科学技术与工程, 2018(6): 285-290.
- [16] ZANGENEH V, SHAJARI M. A cost-sensitive move selection strategy for moving target defense[J]. Computers & Security, 2018, 75: S0167404817302821.
- [17] XIE J, LI M. Software defect location based on cost-sensitive interval distribution optimization[J]. Acta Journal of Software, 2017, 28(11): 3072-3079. (in Chinese)
解铮, 黎铭. 基于代价敏感间隔分布优化的软件缺陷定位[J]. 软件学报, 2017, 28(11): 3072-3079.
- [18] SHI Y W, WANG H J. A cost-sensitive random forest classifier based on a new measure of impureness[J]. Journal of Computer Science, 2017, 44(S2): 98-101. (in Chinese)
师彦文, 王宏杰. 基于新型不纯度度量的代价敏感随机森林分类器[J]. 计算机科学, 2017, 44(S2): 98-101.
- [19] ZHAO P, ZHANG Y, WU M, et al. Adaptive Cost-Sensitive Online Classification[J]. IEEE Transactions on Knowledge & Data Engineering, 2018, PP(99): 1-1.
- [20] VERMA A, HANAWAL M K. Unsupervised cost sensitive predictions with side information[C]// the ACM India Joint International Conference. ACM, 2018: 364-367.
- [21] BACH M, WERNER A. Cost-Sensitive Feature Selection for Class Imbalance Problem[M]// Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology — ISAT 2017. 2018: 182-194.
- [22] LI Q J, ZHAO Y Q, GU Z, et al. Loss function design in cost-sensitive learning[J]. Control Theory & Applications, 2015, 32(5): 689-694. (in Chinese)
李秋洁, 赵亚琴, 顾洲, 等. 代价敏感学习中的损失函数设计[J]. 控制理论与应用, 2015, 32(5): 689-694.
- [23] CHAN P, STOLFO S. Toward scalable learning with non-uniform class and cost distributions[C]// International Conference on Knowledge Discovery & Data Mining. 1998: 164-168.
- [24] ELKAN C. The Foundations of Cost-Sensitive Learning[C]// Seventeenth International Joint Conference on Artificial Intelligence. 2001: 973-978.
- [25] DOMINGOS P. MetaCost: a general method for making classifiers cost-sensitive[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 1999: 155-164.
- [26] LI J, FONG S, MOHAMMED S, et al. Improving the classification performance of biological imbalanced datasets by swarm optimization algorithms [J]. Journal of Supercomputing, 2016, 72(10): 1-21.
- [27] RYU D, JANG J I, BAIK J. A transfer cost-sensitive boosting approach for cross-project defect prediction[J]. Software Quality Journal, 2017, 25(1): 1-38.
- [28] RYU D, JANG J I, BAIK J. A transfer cost-sensitive boosting approach for cross-project defect prediction[J]. Software Quality Journal, 2017, 25(1): 235-272.
- [29] DRUMMOND C, HOLTE R C. Exploiting the Cost(In)sensitivity of Decision Tree Splitting Criteria[C]// Seventeenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc. 2000: 239-246.
- [30] LOMAX S, VADERA S. A survey of cost-sensitive decision tree induction algorithms[J]. Acm Computing Surveys, 2013, 45(2): 1-35.
- [31] KNOLL U, NAKHAEIZADEH G, TAUSEND B. Cost-sensitive pruning of decision trees[C]// Machine Learning: ECML-94, European Conference on Machine Learning. Catania, Italy, DBLP, 1994: 383-386.
- [32] BRADFORD J P, KUNZ C, KOHAVI R, et al. Pruning Decision Trees with Misclassification Costs[M]// Machine Learning: ECML-98. Springer Berlin Heidelberg, 1998: 131-136.
- [33] ZHENG E H, LI P, SONG Z H, et al. Cost-Sensitive Support Vector Machines[J]. Control and Decision, 2006, 21(4): 473-476. (in Chinese)
郑恩辉, 李平, 宋执环, 等. 代价敏感支持向量机[J]. 控制与决策, 2006, 21(4): 473-476.
- [34] ADACHI Y, ONIMURA N, YAMASHITA T, et al. Classification of Imbalanced Documents by Feature Selection[C]// International Conference on Compute and Data Analysis. ACM, 2017: 228-232.
- [35] ZHOU Y H, ZHOU Z H. Cost-sensitive large interval distributed learning machine[J]. Journal of Computer Research and Development, 2016, 53(9): 1964-1970. (in Chinese)
周宇航, 周志华. 代价敏感大间隔分布学习机[J]. 计算机研究与发展, 2016, 53(9): 1964-1970.
- [36] KUKAR M Z, KONONENKO I. Cost-Sensitive Learning with Neural Networks[C]// Proceedings of the 13th European Conference on Artificial Intelligence, Wiley, New York, 1998: 445-449.

- [37] ZHOU Z H, LIU X Y. Training cost-sensitive neural networks with methods addressing the class imbalance problem[J]. IEEE Transactions on Knowledge & Data Engineering, 2006, 18(1): 63-77.
- [38] HUANG L W, JIANG B T, LU S Y, et al. A review of recommendation systems based on deep learning [J]. arxiv:1707.07435, 2018. (in Chinese)
黄立威, 江碧涛, 吕守业, 等. 基于深度学习的推荐系统研究综述[J]. arxiv:1707.07435, 2018.
- [39] HE Y X, SUN S T, NIU F F, et al. An in-depth learning model for emotional semantics enhancement in Weibo emotion analysis [J]. Chinese Journal of Computers, 2017, 40(4): 773-790. (in Chinese)
何炎祥, 孙松涛, 牛菲菲, 等. 用于微博情感分析的一种情感语义增强的深度学习模型[J]. 计算机学报, 2017, 40(4): 773-790.
- [40] CHENG X Q, QI X L, WANG Y Z, et al. A review of big data systems and analysis techniques[J]. Journal of Software, 2014, 25(9): 1889-1908. (in Chinese)
程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述[J]. 软件学报, 2014, 25(9): 1889-1908.
- [41] YAN X F, ZHOU G D. Research on Deep Learning for Natural Language Processing[J]. Journal of Automation, 2016, 42(10): 1445-1465. (in Chinese)
奚雪峰, 周国栋. 面向自然语言处理的深度学习研究[J]. 自动化学报, 2016, 42(10): 1445-1465.
- [42] KAMILARIS A, PRENAFETA-BOLDÚ F X. Deep learning in agriculture: A survey[J]. Computers & Electronics in Agriculture, 2018, 147(1): 70-90.
- [43] HATCHER W G, YU W. A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends[J]. IEEE Access, 2018, PP(99): 1-1.
- [44] MA S L, WUNIRI Q Q G, LI X P. Summary of big data and deep learning [J]. Journal of Intelligent Systems, 2016, 11(6): 728-742. (in Chinese)
马世龙, 乌尼日其其格, 李小平. 大数据与深度学习综述[J]. 智能系统学报, 2016, 11(6): 728-742.
- [45] HINTON G, OSINDERO S, TEH Y. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [46] BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy layer-wise training of deep networks[C]// Proc of the 12th Annual Conference on Neural Information Processing System. 2006: 153-160.
- [47] BENGIO Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
- [48] BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy Layer-wise training of deep networks [C]// Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2007: 153-160.
- [49] YIN B C, WANG W T, WANG L C. A summary of deep learning research [J]. Journal of Beijing University of Technology, 2015, 41(1): 48-59. (in Chinese)
尹宝才, 王文通, 王立春. 深度学习研究综述[J]. 北京工业大学学报, 2015, 41(1): 48-59.
- [50] CHEN W H, AN J Y, LI R F, et al. Summary of deep learning cognitive computing [J]. Automation Journal, 2017, 43(11): 1886-1897. (in Chinese)
陈伟宏, 安吉尧, 李仁发, 等. 深度学习认知计算综述[J]. 自动化学报, 2017, 43(11): 1886-1897.
- [51] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137-1149.
- [52] CHENG B, WEI Y, SHI H, et al. Revisiting RCNN: On Awakening the Classification Power of Faster RCNN[M]// Computer Vision (ECCV 2018). 2018.
- [53] MAHAPATRA D, ROY P K, SEDAI S, et al. Retinal Image Quality Classification Using Saliency Maps and CNNs[J]. Interacting with Computers, 2016, 10(4): 385-399.
- [54] ZHONG Y, FEI F, LIU Y, et al. SatCNN: satellite image dataset classification using agile convolutional neural networks[J]. Remote Sensing Letters, 2017, 8(2): 136-145.
- [55] YUE S. Imbalanced Malware Images Classification: a CNN based Approach[J]. arXiv:1708.08042, 2017.
- [56] WINDRIM L, RAMAKRISHNAN R, MELKUMYAN A, et al. Hyperspectral CNN Classification with Limited Training Samples[J]. arXiv:1611.09007v1, 2016.
- [57] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [58] MNH V, HEES N, GRAVES A, et al. Recurrent Models of Visual Attention[J]. Advances in Neural Information Processing Systems, 2014: 2204-2212.
- [59] SHEN L, ZHANG J. Empirical Evaluation of RNN Architectures on Sentence Classification Task[J]. arXiv:1609.09171, 2016.
- [60] KANG S I, LEE S M. Improvement of Speech/Music Classification Based on RNN in EVS Codec for Hearing Aids[J]. Journal of Rehabilitation Welfare Engineering & Assistive Technology, 2017, 11.
- [61] MUN S, PARK D, KO H. A Novel Discriminative Feature Extraction for Emotion Classification using RNN based Source Separation[J]. IEEE Transactions on Information & Systems, 2017, 100(12): 3041-3044.
- [62] CHENG J, LI P, DING Z, et al. Sentiment Classification of Chinese Microblogging Texts with Global RNN[C]// IEEE International Conference on Data Science in Cyberspace. IEEE, 2017: 653-657.
- [63] WANG J, YANG Y, MAO J, et al. CNN-RNN: A Unified Framework for Multi-label Image Classification[J]. arXiv:1604.04573, 2016.
- [64] LIU F, LI S, ZHANG L, et al. 3DCNN-DQN-RNN: A Deep Reinforcement Learning Framework for Semantic Parsing of Large-Scale 3D Point Clouds[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017: 5679-5688.
- [65] HORI T, WATANABE S, ZHANG Y, et al. Advances in Joint

- CTC-Attention based End-to-End Speech Recognition with a Deep CNN Encoder and RNN-LM[J]. arXiv:1706.02737, 2017.
- [66] XIONG H K, GAO X, LI Y H, et al. Explicit, structured and multimodal deep neural networks[J]. Pattern Recognition & Artificial Intelligence, 2018, 31(1): 1-11. (in Chinese)
熊红凯, 高星, 李劭辉, 等. 可解释化、结构化、多模态化的深度神经网络[J]. 模式识别与人工智能, 2018, 31(1): 1-11.
- [67] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// Proceedings of the 2014 Conference on Advances in Neural Information Processing Systems 27. Montreal, Canada: Curran Associates, Inc., 2014: 2672-2680.
- [68] WANG K F, GOU C, DUAN Y J, et al. Generative Adversarial Networks: The State of the Art and Beyond[J]. ACTA AUTOMATICA SINICA, 2017, 43(3): 321-332. (in Chinese)
王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望[J]. 自动化学报, 2017, 43(3): 321-332.
- [69] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[J]. arXiv:1701.07875, 2017.
- [70] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets[C]// Proceedings of the 2016 Neural Information Processing Systems. 2016.
- [71] LI Y B, ZHOU S M. A New Loss Function[J]. Journal of Wuhan University of Technology, 1999, 21(4): 86-87. (in Chinese)
李跃波, 周树民. 一种新的损失函数[J]. 武汉理工大学学报, 1999, 21(4): 86-87.
- [72] WANG S, LIU W, WU J, et al. Training deep neural networks on imbalanced data sets[C]// International Joint Conference on Neural Networks. IEEE, 2016: 4368-4374.
- [73] CHENG D, GONG Y, ZHOU S, et al. Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016: 1335-1344.
- [74] ZHANG X L. Speech separation by cost-sensitive deep learning[C]// APSIPA. 2017: 159-162.
- [75] JIANG J, LIU X, ZHANG K, et al. Automatic diagnosis of imbalanced ophthalmic images using a cost-sensitive deep convolutional neural network[J]. Biomedical Engineering Online, 2017, 16(1): 132.
- [76] HUANG C, LI Y, CHEN C L, et al. Learning Deep Representation for Imbalanced Classification[C]// Computer Vision and Pattern Recognition. IEEE, 2016: 5375-5384.
- [77] KHAN S H, HAYAT M, BENNAMOUN M, et al. Cost-Sensitive Learning of Deep Feature Representations From Imbalanced Data[J]. IEEE Transactions on Neural Networks & Learning Systems, 2015, PP(99): 1-15.
- [78] YUAN C, LIN H T, YANG S W. Cost-aware Pre-training for Multiclass Cost-sensitive Deep Learning[J]. arXiv:1511.09337, 2015.
- [79] YUAN C, LIN H T. Cost-Sensitive Deep Learning with Layer-Wise Cost Estimation[J]. arXiv:1611.05134, 2016.
- [80] TU H H, LIN H T. One-sided support vector regression for multiclass cost-sensitive classification[C]// International Conference on International Conference on Machine Learning. Omnipress, 2010: 1095-1102.
- [81] LI H, ZHOU X, HUANG B, et al. Cost-Sensitive Three-Way Decision: A Sequential Strategy[M]// Rough Sets and Knowledge Technology. Springer Berlin Heidelberg, 2013: 325-337.
- [82] PANG J B, LIN H H, DUAN L J, et al. Accelerate convolutional neural networks for binary classification: a cascading cost-sensitive feature approach[C]// Proceedings of the 13th European Conference on Artificial Intelligence (ECAI-98). 2016.
- [83] LEE Y J, MANGASARIAN O L. SSVM: A Smooth Support Vector Machine for Classification[J]. Computational Optimization & Applications, 2001, 20(1): 5-22.
- [84] HINTON G E, OSINDERO S, TEH Y. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [85] VINCENT P. A connection between score matching and denoising autoencoders[M]. MIT Press, 2011.
- [86] RAJ V, MAGG S, WERMTER S. Towards Effective Classification of Imbalanced Data with Convolutional Neural Networks[M]// Artificial Neural Networks in Pattern Recognition. Springer International Publishing, 2016: 150-162.
- [87] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv:1409.1556, 2014.
- [88] FIORE U, SANTIS A D, PERLA F, et al. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection[J]. Information Sciences, 2017: S0020025517311519.
- [89] GENG Y, LUO X. Cost-Sensitive Convolution based Neural Networks for Imbalanced Time-Series Classification[J]. arXiv: 1801.04396 2018.